

## RAW SEQUENCE LISTING

The Biotechnology Systems Branch of the Scientific and Technical Information Center (STIC) no errors detected.

Application Serial Number: 09/830,691A  
Source: IFWO  
Date Processed by STIC: 1-6-4

# ***ENTERED***

RE-RUN



IFWO

## RAW SEQUENCE LISTING

DATE: 01/06/2005

PATENT APPLICATION: US/09/830,691A

TIME: 14:49:08

Input Set : N:\Cr4\Refhold\09\_folder\I830691A.raw

Output Set: N:\CRF4\01062005\I830691A.raw

1 <110> APPLICANT: Choi, Eui-Sung  
2 Rhee, Sang-Ki  
3 Sohn, Jung-Hoon  
4 Park, Soo-Dong  
5 Lee, Yoon-Hyoung  
6 Lee, Seung-Jae  
7 Jang, Jae-Kweon  
8 Choi, Seok-Keun  
9 Son, Young-Rok  
10 <120> TITLE OF INVENTION: VECTOR FOR THE TRANSFORMATION OF PHAFFIA  
11 RHODOZYMA AND PROCESS OF TRANSFORMATION THEREBY  
12 <130> FILE REFERENCE: 118.12-US-WO  
C--> 13 <140> CURRENT APPLICATION NUMBER: US/09/830,691A  
14 <141> CURRENT FILING DATE: 2001-04-26  
15 <150> PRIOR APPLICATION NUMBER: KR 1998/46547  
16 <151> PRIOR FILING DATE: 1998-10-31  
17 <150> PRIOR APPLICATION NUMBER: PCT/KR99/00265  
18 <151> PRIOR FILING DATE: 1999-05-29  
19 <160> NUMBER OF SEQ ID NOS: 20  
20 <170> SOFTWARE: FastSEQ for Windows Version 4.0  
22 <210> SEQ ID NO: 1  
23 <211> LENGTH: 1223  
24 <212> TYPE: DNA  
25 <213> ORGANISM: Phaffia rhodozyma  
26 <400> SEQUENCE: 1  
27 atggtcaacg ttcccaagac tcgacgtgag ttatagcaat ttcaacaact ctccagacga 60  
28 caaatattcc agtgcacga aagagtttgt ggataaacgc gacagtttca agggaaagag 120  
29 tcgatggaca gatttggaag acttagccgg tcaaggaaact tggggatcac gtggcgagg 180  
30 actcatcaga agaagtcggg atttgtttga tcatagtggg atcaagacaa actggaggat 240  
31 atggctcgcc ttggaaggga atctcgggcc tggattcgag gatccgaaaag ttgtacgtat 300  
32 ggaaaagctt acacggcttg gatttattat ctttcatagg aacctactgc aagggttaagg 360  
33 cttgcaagaa gcacacgtaa gtcgcttata ctctccactc ttcatggca tattgtcaac 420  
34 gactggacaa cgcgtccgtt ttgaaacaag tgacttacct gtgaaatttg attctacacc 480  
35 tgtatttagc cctcacaagg tacatatcac atcctcccac cccaccctgc ccaacttctt 540  
36 cagttcatct tgctctcggt ttccacattc cctgatgacc tccttgtagt ttctttgcca 600  
37 acgtttgttt ctgtttctgt aggtgaccca gtacaagaag ggaaaggact ccatottcgc 660  
38 ccagggaaaag cgacgatacg accgaaagaa gtocggttac ggaggtcaga ccaagcccg 720  
39 ttccacaag aaggctaaga ccaccaagaa ggtcgctcct cgattggcgg tatttttgtt 780  
40 tattttgaat tctttttgtg tatgcagact ttgatgatt atgctcctct gtcgtttttt 840  
41 ctottcaaac agagtgtctc gtctgcagtt cgttcttcct tccaaccaa acttcaacta 900  
42 cagacatcat aaacagacat cttacttcgg tgttctctct tttttccgc agagtacaag 960  
43 atgcagatga ccctcaagcg atgcaagcac ttcgagcttg gaggagacaa gaagaccaag 1020  
44 ggttcgtctt ttgtccatat attctctggt tcacttctta tgttcctaac gtacttggtt 1080

## RAW SEQUENCE LISTING

DATE: 01/06/2005

PATENT APPLICATION: US/09/830,691A

TIME: 14:49:08

Input Set : N:\Crf4\Refhold\09\_folder\I830691A.raw

Output Set: N:\CRF4\01062005\I830691A.raw

```

45      cctttttggt tgggatgttg ttctatcggt tgggtgttttc tttcttttgg atgcattatc 1140
46      atttatcggt ttggactgtt ttcctctgct cggtttctttc tccctctgtac ttgtgtctct 1200
47      caggagccgc catctctttc taa                                     1223
49 <210> SEQ ID NO: 2
50 <211> LENGTH: 350
51 <212> TYPE: DNA
52 <213> ORGANISM: Phaffia rhodozyma
53 <220> FEATURE:
54 <221> NAME/KEY: CDS
55 <222> LOCATION: (30)...(347)
56 <400> SEQUENCE: 2
57      cccttcaagt ctctgtctcaa tcagtcaag atg gtc aac gtt ccc aag act cga 53
58      Met Val Asn Val Pro Lys Thr Arg
59      1 5
60      cga acc tac tgc aag ggt aag gct tgc aag aag cac acc cct cac aag 101
61      Arg Thr Tyr Cys Lys Gly Lys Ala Cys Lys Lys His Thr Pro His Lys
62      10 15 20
63      gtg acc cag tac aag aag gga aag gac tcc atc ttc gcc cag gga aag 149
64      Val Thr Gln Tyr Lys Lys Gly Lys Asp Ser Ile Phe Ala Gln Gly Lys
65      25 30 35 40
66      cga cga tac gac cga aag cag tcc ggt tac gga ggt cag acc aag ccc 197
67      Arg Arg Tyr Asp Arg Lys Gln Ser Gly Tyr Gly Gly Gln Thr Lys Pro
68      45 50 55
69      gtt ttc cac aag aag gct aag acc acc aag aag gtc gtc ctt cga ttg 245
70      Val Phe His Lys Lys Ala Lys Thr Lys Lys Val Val Leu Arg Leu
71      60 65 70
72      gag tgc tcc gtc tgc aag tac aag atg cag atg acc ctc aag cga tgc 293
73      Glu Cys Ser Val Cys Lys Tyr Lys Met Gln Met Thr Leu Lys Arg Cys
74      75 80 85
75      aag cac ttc gag ctt gga gga gac aag aag acc aag gga gcc gcc atc 341
76      Lys His Phe Glu Leu Gly Gly Asp Lys Lys Thr Lys Gly Ala Ala Ile
77      90 95 100
78      tct ttc taa                                     350
79      Ser Phe
80      105
82 <210> SEQ ID NO: 3
83 <211> LENGTH: 106
84 <212> TYPE: PRT
85 <213> ORGANISM: Phaffia rhodozyma
86 <400> SEQUENCE: 3
87      Met Val Asn Val Pro Lys Thr Arg Arg Thr Tyr Cys Lys Gly Lys Ala
88      1 5 10 15
89      Cys Lys Lys His Thr Pro His Lys Val Thr Gln Tyr Lys Lys Gly Lys
90      20 25 30
91      Asp Ser Ile Phe Ala Gln Gly Lys Arg Arg Tyr Asp Arg Lys Gln Ser
92      35 40 45
93      Gly Tyr Gly Gly Gln Thr Lys Pro Val Phe His Lys Lys Ala Lys Thr
94      50 55 60
95      Thr Lys Lys Val Val Leu Arg Leu Glu Cys Ser Val Cys Lys Tyr Lys

```

## RAW SEQUENCE LISTING

DATE: 01/06/2005

PATENT APPLICATION: US/09/830,691A

TIME: 14:49:08

Input Set : N:\Crf4\Refhold\09\_folder\I830691A.raw

Output Set: N:\CRF4\01062005\I830691A.raw

```

96      65      70      75      80
97      Met Gln Met Thr Leu Lys Arg Cys Lys His Phe Glu Leu Gly Gly Asp
98      85      90      95
99      Lys Lys Thr Lys Gly Ala Ala Ile Ser Phe
100     100     105
102 <210> SEQ ID NO: 4
103 <211> LENGTH: 741
104 <212> TYPE: DNA
105 <213> ORGANISM: Phaffia rhodozyma
106 <220> FEATURE:
107 <221> NAME/KEY: misc_feature
108 <222> LOCATION: (0)...(0)
109 <223> OTHER INFORMATION: n=a, t, c, or g
110 <400> SEQUENCE: 4
111      ctcgagtggg cggtggcaat ggcattcgtg tcgttggtgc tcactcgcaa cccaagcagt 60
112      cgcttaccgc gggtagcctc cgggtgggcg cgatgatttg tgggtgtggat tccttcctta 120
113      tgggttagaac gacgcgcaac caatcattcg gagaaccgct ccgttgttagc cgaccagtct 180
114      gattgatcaa catgccagca cgtcctccgg gacggagact ggcggggatc gtacctcatc 240
115      tgggaatcgct ggtcaatgg tagtagtctt cagcatcggc catgagggca gtctaggtgg 300
W--> 116      gttcgcctgc cgaagactgt gtgagtgtgc tganaactaa ttgagtaccg ggggataagg 360
117      caaggcgtgt ntggttgccg gtggctgtga gcgagtttgc tgcaaagcga ttcaatgcac 420
118      cccggcttgg ccagcgcgct gcgtcacgaa acacactaaa cggttgacgc cataaagtaa 480
119      taacacactc aagtttgtgg tcccgggtgg gcctctgtgc ctgcgtggga cccgacggga 540
120      gaggaaaaacg ttctgtggcc ctctcctctg tggatagtta cctggttgat cctgccagta 600
121      gtcatatgct tgtctcaaag attaagccat gcattgtctaa gtataaacia attcactctg 660
122      tgaaactgcg aatggctcat taaatcagtt atagtttatt tgatgggtacc ttgctacatg 720
123      gataactgtg gtaattctag a                                     741
125 <210> SEQ ID NO: 5
126 <211> LENGTH: 23
127 <212> TYPE: DNA
128 <213> ORGANISM: Artificial Sequence
129 <220> FEATURE:
130 <223> OTHER INFORMATION: CYH1, a PCR primer for the cloning of L41 genomic
131      DNA fragment
132 <220> FEATURE:
133 <221> NAME/KEY: misc_feature
134 <222> LOCATION: (0)...(0)
135 <223> OTHER INFORMATION: n=a, t, c, or g
136 <400> SEQUENCE: 5
W--> 137      cgcgtagtta aygtncnaa rac                                     23
139 <210> SEQ ID NO: 6
140 <211> LENGTH: 25
141 <212> TYPE: DNA
142 <213> ORGANISM: Artificial Sequence
143 <220> FEATURE:
144 <223> OTHER INFORMATION: CYH3, a PCR primer for the cloning of L41 genomic
145      DNA fragment
146 <400> SEQUENCE: 6
147      cccgggtytt ggcytttytr tgraa                                     25
```

## RAW SEQUENCE LISTING

DATE: 01/06/2005

PATENT APPLICATION: US/09/830,691A

TIME: 14:49:08

Input Set : N:\Cr4\Refhold\09\_folder\I830691A.raw

Output Set: N:\CRF4\01062005\I830691A.raw

149 <210> SEQ ID NO: 7  
150 <211> LENGTH: 24  
151 <212> TYPE: DNA  
152 <213> ORGANISM: Artificial Sequence  
153 <220> FEATURE:  
154 <223> OTHER INFORMATION: 3' RACE primer  
155 <400> SEQUENCE: 7  
156 ggtcagacca agcaagtttt tcac 24  
158 <210> SEQ ID NO: 8  
159 <211> LENGTH: 24  
160 <212> TYPE: DNA  
161 <213> ORGANISM: Artificial Sequence  
162 <220> FEATURE:  
163 <223> OTHER INFORMATION: 5' RACE primer  
164 <400> SEQUENCE: 8  
165 gtgaaaaact tgcttggtct gacc 24  
167 <210> SEQ ID NO: 9  
168 <211> LENGTH: 24  
169 <212> TYPE: DNA  
170 <213> ORGANISM: Artificial Sequence  
171 <220> FEATURE:  
172 <223> OTHER INFORMATION: sense primer for the mutagenesis of L41 gene  
173 <400> SEQUENCE: 9  
174 ggtcagacca agcaagtttt tcac 24  
176 <210> SEQ ID NO: 10  
177 <211> LENGTH: 24  
178 <212> TYPE: DNA  
179 <213> ORGANISM: Artificial Sequence  
180 <220> FEATURE:  
181 <223> OTHER INFORMATION: antisense primer for the mutagenesis of L41 gene  
182 <400> SEQUENCE: 10  
183 gtgaaaaact tgcttggtct gacc 24  
185 <210> SEQ ID NO: 11  
186 <211> LENGTH: 20  
187 <212> TYPE: DNA  
188 <213> ORGANISM: Artificial Sequence  
189 <220> FEATURE:  
190 <223> OTHER INFORMATION: a PCR primer corresponding to 18S rDNA  
191 <400> SEQUENCE: 11  
192 tcctagtaag cgcaagtcac 20  
194 <210> SEQ ID NO: 12  
195 <211> LENGTH: 20  
196 <212> TYPE: DNA  
197 <213> ORGANISM: Artificial Sequence  
198 <220> FEATURE:  
199 <223> OTHER INFORMATION: a PCR primer corresponding to 18S rDNA  
200 <400> SEQUENCE: 12  
201 ttccggccaag gaaagaaact 20  
203 <210> SEQ ID NO: 13

## RAW SEQUENCE LISTING

DATE: 01/06/2005

PATENT APPLICATION: US/09/830,691A

TIME: 14:49:08

Input Set : N:\Crf4\Reifhold\09\_folder\I830691A.raw

Output Set: N:\CRF4\01062005\I830691A.raw

```
204 <211> LENGTH: 20
205 <212> TYPE: DNA
206 <213> ORGANISM: Artificial Sequence
207 <220> FEATURE:
208 <223> OTHER INFORMATION: a PCR primer corresponding to 28S rDNA
209 <400> SEQUENCE: 13
210      aatcgggatta tccggagcta                                20
212 <210> SEQ ID NO: 14
213 <211> LENGTH: 20
214 <212> TYPE: DNA
215 <213> ORGANISM: Artificial Sequence
216 <220> FEATURE:
217 <223> OTHER INFORMATION: a PCR primer corresponding to 28S rDNA
218 <400> SEQUENCE: 14
219      gctataaacac atccggagat                                20
221 <210> SEQ ID NO: 15
222 <211> LENGTH: 2192
223 <212> TYPE: DNA
224 <213> ORGANISM: Phaffia rhodozyma
225 <400> SEQUENCE: 15
226      aagagctatt tgaatgacga ccacaagagt gacgatcata ttgagcatag tataccaaag 60
227      gccaaagaggo tgtgtggtgt tctatgagtg gccttgatta tgtgttacat aaataaactg 120
228      atctcaatttt ttcaaatact tgcoaacact ttoatatatt cacaccaaaa aaagtcagat 180
229      tggcccacaa agtcagatac acgctcgatc gtgcagcgggt tcaagcactt tgtcaggcga 240
230      aagaaaggcc acagcaccac ccttcaagtc tcgtctcaat caggttcgtc tagctttttg 300
231      tgtgcaagga tttaccgtct tgatggattt gttcgttgaa agagaggaaa gaacatgctg 360
232      aactgacgaa agtgtgaaca aaaaattgtg attttttcat tgtgtttcgc tgggtctcctt 420
233      gctgggttgg gttggatcgg attttatctt tgtgttggat ggaaaaccct gaatgttctt 480
234      ttcttggaca tcttctaacc tcgacaaaaa gattcattcc tccgtactgc totggttctg 540
235      cctttttgaa tcgcattgat aaattcttcc ctgcgaacgt tcgatcaatc tccgtcaaac 600
236      ttatcatcca aaaatctctt ctgcactgcc gccttgctcc ttttcttctt tctttcctta 660
237      atccgctttc gactaccctc cttctcttca cactcatagt caagatgggc aacgttccca 720
238      agactcgacg tgagttatag caatttcaac aactctccag acgacaaaata ttccagtgcg 780
239      tcgaaagagt ttgtggataa acgcgacagt ttoaagggaa agagtcgatg gacagatttg 840
240      gaagacttag ccggtcaagg aacttgggga tcacgtggcg gaggactcat cagaagaagt 900
241      cgggatttgt ttgatcatag tgggatcaag acaaaactga ggatatggct cgccttggaa 960
242      gggaaatctcc ggcttggatt cgaggatccg aaagttgtac gtatggaaaa gcttacacgg 1020
243      cttggattta ttatctttca taggaacctc ctgcaagggt aaggcttgca agaagcacac 1080
244      gtaagtcgct tatcctctcc actctttcat ggcattattg caacgactgg acaacgcgtc 1140
245      cgttttgaaa caagtgactt acctgtgaaa tttgattcta cactgtatt tagccctcac 1200
246      aaggtacata tcacatcctc ccaccccaac ctgcccaact tcttcagttc atcttgotct 1260
247      cggtttccac attccttgat gacctccttg tatgttcttt gcgaacgttt gtttctgttt 1320
248      ctgtaggtga cccagtacaa gaagggaag gactccatct tcgcccaggg aaagcgacga 1380
249      tacgaccgaa agcagtcogg ttacggaggt cagaccaagc ccgtttttca caagaaggct 1440
250      aagaccacca agaaggtcgt ccttcgattg ggtacgtttt tgtttatttt gaattctttt 1500
251      tgtgtatgca gacttttgat gattatgctc ctctgtcgtt ttttctcttc aaacagagtg 1560
252      ctccgtctgc agttcgttct tccttccaac caaaacttca actacagaca tcataaacag 1620
253      acatcttact tcggtgttct ctcttttttt ccgcagagta caagatgcag atgaccctca 1680
254      agcgatgcaa gcacttcgag cttggaggag acaagaagac caagggttct tcttttgtcc 1740
```

RAW SEQUENCE LISTING ERROR SUMMARY      DATE: 01/06/2005  
PATENT APPLICATION: US/09/830,691A      TIME: 14:49:09

Input Set : N:\CrF4\Refhold\09\_folder\I830691A.raw  
Output Set: N:\CRF4\01062005\I830691A.raw

Please Note:

Use of n and/or Xaa have been detected in the Sequence Listing. Please review the Sequence Listing to ensure that a corresponding explanation is presented in the <220> to <223> fields of each sequence which presents at least one n or Xaa.

Seq#:4; N Pos. 334,371  
Seq#:5; N Pos. 15,18

VERIFICATION SUMMARY

DATE: 01/06/2005

PATENT APPLICATION: US/09/830,691A

TIME: 14:49:09

Input Set : N:\Crf4\Refhold\09\_folder\I830691A.raw

Output Set: N:\CRF4\01062005\I830691A.raw

L:13 M:270 C: Current Application Number differs, Wrong Format

L:116 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:4 after pos.:300

M:341 Repeated in SeqNo=4

L:137 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:5 after pos.:0